ENCYCLOPEDIA OF THE HUMAN GENOME
2000

Molecular genetics: robotics and automation

Reference code:351

Trevor Hawkins, PhD.
Director
U.S. D.O.E. Joint Genome Institute
2800 Mitchell Drive
Walnut Creek, CA 94598 U.S.A.
tlhawkins@lbl.gov

Chris Elkin, M.S.
Manager – Advanced Sequencing Technology
U.S. D.O.E. Joint Genome Institute
2800 Mitchell Drive
Walnut Creek, CA 94598 U.S.A.
CJElkin@lbl.gov

Martin Pollard, M.S.
Manager – Instrumentation
U.S. D.O.E. Joint Genome Institute
2800 Mitchell Drive
Walnut Creek, CA 94598 U.S.A.
MJPollard@lbl.gov

Molecular genetics: robotics and automation

Hawkins, Trevor

Chris Elkin, Martin Pollard, Trevor Hawkins

Human Genome Project, Automation, Robotics, DNA Sequencing

INTRODUCTION

Automation has been the backbone of the human genome project and the foundation of the biotech/ genomics industry. In our quest to turn lab methods into industrial scale processes we have turned largely to robotic solutions in tandem with new biochemical approaches. The past few years has seen many changes in the types of automation used as new methods are developed and the scale of the processes increased. It is also the case that the availability of automation has changed dramatically over the past 5 years as dozens of companies have started to supply custom and off the shelf systems.

In this report we describe some of the latest automated systems used within the human genome project in terms of sample processing and preparation (the 'front end') and DNA sequencing and detection devices.

SAMPLE PROCESSING AND PREPARATION

Sample processing automation in genomics is designed to assist in the processing of samples prior to loading in DNA sequencing instruments. This generally includes the process steps: creating a library of cloned DNA from the genome, isolating and purifying the DNA, and attaching the labeled fluorescent tags to the DNA prior to electrophoresis and detection. During the course of the Human Genome Project a great deal of commercial automation has been developed to address these most common process needs. Some of the instruments will only perform a single, specialized task while others are flexible programmable systems that can be programmed to perform multiple steps under user control. Industry standards for microtiter plates have resulted in instruments from all manufacturers to be generally compatible with each other. While all genomic sequencing strategies share common types of steps the specific protocol details at each genome center or company require instruments to be individually configurable to some degree. The common major automation platforms in use today are described below.

DNA Shearing

DNA fragments to be used to create sub clone libraries can be created using enzymatic, ultrasonic, or mechanical shearing methods. One automated mechanical shearing instrument, the Genemachines HydroShear (www.genemachines.com) shears DNA via hydrodynamic shearing by pumping the DNA sample through an orifice plate. As a pump pushes the DNA through a small hole in the plate, the velocity of the sample is increased locally to stretch and fragment the DNA. The velocity of the pump and the size of the orifice tune the average DNA fragment size. Typically genomic or large DNA is sheared into 2,000bp average fragments which are then end repaired and used to clone into sub clone vectors to produce libraries.

Colony/Plaque Pickers

Colony picking was one of the first sample processing tasks to be automated dating back to systems developed by academic groups in the early 1980's. All colony pickers operate on the same basic design; randomly arrayed colonies/plaques are presented to the picker on agar plates. A CCD camera imaging system is then used to guide the picking needles by creating a digital image of the colony locations and converting these to mechanical picking coordinates. The needles are then moved to the physical coordinates and a small number of cells picked onto the needle tips. A destination microtiter plate, either 96 or 384 wells are filled with growth media, and then inoculated with the colonies. Finally, the picking needles are cleaned using combinations of ultrasonic baths, mechanical scrubbing and heat drying to be then reused in the next picking cycle. The imaging software selects colonies based on user defined criteria used to select single isolated colonies. Throughputs of these devices vary depending on colony density and imaging criteria, the length of the tip cleaning procedure, whether the machines are manually loaded or are equipped with autoloaders. Throughputs should be evaluated based on the number of output samples/hour and can range from 1,000 to 4,000 in most currently available systems. Examples of systems currently available are: Genetix Qbot & Qpix (www.genetix.com), Genemachines Gel-2-Well (www.genemachines.com) and, Genomic Solutions Flexys (www.genomicsolutions.com).

Thermal Cyclers

Thermal cyclers run programmed temperature cycling of samples for processes such as polymer chain reaction (PCR) and cycle sequencing reactions. The most popular thermal cyclers operate using peltier effect devices that can heat or cool electrically such as the Perkin Elmer Geneamp 9700 (www.appliedbiosystems.com), MJ Research Tetrad (www.mjresearch.com) Thermal Cycler. The typical operating range for these thermal cyclers is 4 degrees C to 100 degrees C with ramp times in the range or 1-2 degrees C per second. The instruments can be programmed for any sequence of temperatures, ramp speeds and dwell times. These thermal cyclers require the samples to be processed in thin walled microtiter plates with either 96 or 384 well plates to ensure good rapid heat transfer from the temperature reservoir and the sample using heated lids to prevent evaporation of the sample. Water based instruments are also available for some applications though they usually only have three temperature reservoirs and are typically used for high throughput highly specialized applications. For use of small reaction volumes, less than 5 microliters, capillary tubes can be used in place of the thin walled microtiter plates. These capillary tubes can be heated and cooled very fast and in the case of the Idaho Technology RapidCycler(www.idahotech.com) can provide very rapid cycle times and significant reductions in reaction volumes.

Automated Pipetting/Dispensing

Pipetting and dispensing are some of the most common automated processes in genomics. Pipetting and dispensing instruments are available in 1, 2, 4, 8, 96, and 384 channel systems, i.e. 384 samples are moved together in the 384-channel system. Genomics applications generally use standard microtiter plates such as 96 or 384 wells, although even 1536 well plates are now being introduced. Typical pipettor systems, such as the Tecan Genesis RSP, Packard Instruments Multiprobe and PlateTrak, Matrix PlateMatePlus, Robbins Hydra Microdispenser almost always use some form of syringe mechanism to aspirate and dispense liquids. Instruments are often available in fixed tip arrays of 8, 96, or 384 channels for dispensing or pipetting. All channels operate in unison in these systems. Some pipettor systems are designed with individually programmable pipettors than can randomly access individual wells of plates. This feature is especially useful in rearraying specific wells from many plates into a single working plate.

Pipetting robot systems are available in a variety of designs. Some systems have stationary pipetting heads and move the plates to the pipetting heads using grippers, plate shuttles, or conveyor belts. Others move the heads within the work envelope to access an array of plates spread out on the work surface. While some systems have integral plate moving mechanisms and can be integrated with external robot arms.

Liquid handling systems are generally designed to operate in the 2-200 microliter volume range. Successful applications are very much dependent on pipetting technique and accuracy requirements. The pipetting of 80 microliters of growth media for use on the colony and plaque pickers described above doesn't generally require high accuracy. However, the addition of sequence chemistry reagents at 1-2 microliters does require high accuracy (better than 5% variance) and high reliability (many hundreds of dispensing steps per day). Each application therefore requires decisions to be made about the speed and approach made. Dispensing volumes less than 2 microliters requires careful techniques and modifications to approaches. It is relatively difficult to dispense 1 microliter of liquid into the dry well of a standard polypropylene microtiter plate since liquid viscosity, nozzle material, plastic plate material, speed requirements all factor into protocol development. Future systems are being developed using ink-jet style technologies to dispense nanoliter volumes into specialized plates and reaction chambers. With these types of systems come additional issues such as evaporation, cross contamination and accuracy.

Plate moving systems

Plate moving systems, as defined here, consist of a mechanical arm and gripper or a conveyor belt integrated with several peripheral modules. In recent years, the use of such plate systems has been very popular, such as the dedicated track systems used at the Whitehead Institute (www.wi.mit.edu) and the articulated robotic arm systems employed at the US DOE Joint Genome Institute (www.jgi.doe.gov). Robot arms can be as simple as a cylindrical robot with just vertical and rotational motion plus a gripper to sophisticated articulated arms with as many as six axis of motion such as those developed by CRS robotics (www.crsrobotics.com). Robot arms come in many configurations such as polar, cartesian (xyz), gantry, cantilever for general applications as wells as highly custom designs for specialized applications. Conveyor belts are well known inexpensive systems for moving samples between peripherals in

a prescribed sequence and have been used to great effect by Packard Biosciences Inc (www.packardbioscience.com). Grippers can be designed to grasp almost any type of object and can be controlled with motors or pneumatics. An essential element of any robot system is the control software. The software must allow the user to teach the robot where and how to move to each point in the system. The software must also operate each of the peripheral devices such as pipettors, plate washers, plate shakers and to place plates in static stations such as incubators or magnet stations. An important requirement of the software is to control the arm and peripherals so that several operations can proceed in parallel to increase throughput.

There are several advantages to robot systems. Increased productivity can be achieved because an operator can operate several robots and often still have time to perform other laboratory tasks during a batch process. Training costs are reduced because it is often easier to train an operator to use the robot than to perform the protocol manually. Protocol consistency is improved due to the use of an invariant control program, which can report and log errors.

Integrated robot systems have a high initial cost and are most suited for high throughput production processes. However, since plates are moved from module to module it is often the case that methods can be worked out manually with the modules and then automated with the arm or belt system.

DNA SEQUENCING AND DETECTION DEVICES

Modern sequencing instruments have seen major advancements since their introduction over a decade ago. Based on a manual method devised by Frederick Sanger in 1974, these instruments now produce over sixty thousand bases of sequence information in just two hours. While planned upgrades such as the 384 capillary systems will increase this capacity four fold, several new technologies may provide this data in just minutes or seconds. We will start by reviewing the designs of mainstream high-throughput sequencing instruments. Then we will delve into the theoretical technologies that may lead to the next generation of DNA sequencers.

Initial Slab Gel Sequencing Instruments

Modern versions of Fredrick Sanger's original technique generate fluorescently labeled DNA fragments which are read in an automated fashion with electronic data capture. All possible fragment sizes from 25 bases up to ~1000 bases are produced by the sequencing chemistry with each fragment containing a single fluorescent label that corresponds to the last DNA letter of that fragment. By size fractionating these fragments and matching the color to the size, we can easily discern the original DNA sequence.

The first automated DNA sequencers such as the Applied Biosystems 373 and Amersham Pharmacia ALF were introduced in the early 1990's as slab-gel electrophoresis instruments that utilized a thin (400 μm) cross-linked polyacrylamide gel sandwiched between two glass plates. Dye labeled sequence fragments, initially

primer labeled and now terminator labeled, were manually loaded to the top of the vertical gel assembly and an electric field was applied. As the negatively charged DNA fragments migrated through the gel, they were size fractionated by the polyacrylamide gel. Near the bottom of the gel assembly, separated fragments were automatically excited with a scanning argon laser and detected by a CCD camera. While these instruments automated several steps, slab-gel setup and breakdown were still manual and quite laborious. These issues were addressed in the newer capillary systems.

Capillary Sequencers

Modern capillary systems are currently the most robust and fastest sequencing systems available. These instruments utilize the same basic electrophoresis and detection technology as slab-gel systems with several added advantages. First, each sample is separated and analyzed in an individual capillary. This eliminates sample tracking issues, sample bleed over and gel uniformity problems associated with a slab gel. It is also the case that heat dissipation from a capillary is much more efficient due to its higher surface area-to-volume ratio and this allows for higher run voltages that separate DNA fragments significantly faster than slab gel systems.

The introduction of commercial capillary instruments such as the Applied Biosystems 3700 (www.appliedbiosystems.com) and the Amersham Pharmacia Biotech MegaBACE 1000 (www.apbiotech.com) have automated two key steps; gel preparation and sample loading. Slab gel setup and breakdown were replaced with a linear polyacrylamide gel (LPA) matrix in the case of the MegaBACE and a cross-linked polymer (POP6) in the case of the ABI 3700 that could be forced into and out of the capillary. These replicable gels reduced technician related errors since the new gel required no polymerization and was automatically cleaned out of the capillaries as new gel was injected into the capillaries. Electrokinetic sample injection was another key automation development. With slab gel instruments, samples were hand pipetted on top of the gels leading to misloading and contamination of samples. Electrokinetic injection automated sample loading by passing a negative current through the liquid sample forcing the negatively charged DNA fragments to migrate into the capillary.

Electrokinetic injection and the separation matrices do have drawbacks that are a significant source of variability in sequencing quality. During injection, smaller molecules diffuse faster through the sample than larger molecules. Small charged molecules like salt and dye effectively compete with labeled DNA fragments for injection into the capillary. This lowers the signal strength by inhibiting injection of labeled DNA fragments. Another issue is the injection of large DNA molecules. LPA contains pores that are easily clogged by large DNA molecules like plasmid and genomic DNA. Both of these issues place very high requirements on front-end sample purity and consistency.

A typically capillary instrument produces about 30,000 bases in one hour, a five-fold increase over the original slab sequencers. Further increases of up to twenty-fold will be attained using higher density arrays and new gel formulations. One high-density system due for commercial release in late 2001 is a 384-capillary sequencer that utilizes the same optical and electrophoresis technologies employed in previous systems. Scientists are bundling four times as many capillaries into the MegaBACE sequencer as well as upgrading the optical system to maintain scanning speeds. Where it took nine years to develop the first commercial 96-capillary sequencer, the MegaBACE 4000 DNA sequencer will be completed in just two years.

Future Sequencing Systems

Future sequencing instrument designs will increasingly incorporate more of the upstream processes for total system integration. While extension of electrophoresis technologies like micro-channel-devices promise only a four fold increase in capacity over capillary systems, they offer the ability to etch an entire front end processes onto a single device. This will result in increased process stability and the capability to work with sub-nanoliter volumes.

Theoretical technologies like nanopores and atomic force microscopy could provide the throughputs required for the total genomic analysis of many species. Estimated sequencing speeds of up to 10,000 bases per second could sequence a human genome in a single day. The following is an introduction to these technologies and interested readers are referred to more detailed reviews.

Micro-Channel Devices

Several research groups first described the use of micro-channel devices for DNA sequencing in the early 1990's. These integrated microfluidic devices are similar to capillary instruments except that samples pass through enclosed channels etched into circular, eight-inch diameter, glass or plastic wafers. Advantages include the potential for a total analysis system (known as "lab on a chip"), shorter separation times, inexpensive fabrication and miniaturization for lower sample volumes and increased array density.

To compete with capillary arrays, ninety-six or more channels must be etched into each chip. One limiting issue is that single base resolution requires twenty centimeter long separation channels while current manufacturing technology limits wafer diameters to twenty centimeters. Since the channels can only use half the diameter to eliminate channel crossing, other solutions are required. Scientists at the University of California Berkeley are investigating the use turns to effectively increase the path length of the channels. Further miniaturization of channel space is also possible with new injector designs that greatly improve base resolution. Samples are injected as a small thin plug across the main electrophoresis channel instead of being directly injected into the capillary. With cross-injection, as it is called, the sample enters the

main electrophoresis channel from a sample channel and then flows out of the main channel into a waste channel. Impurities flowing ahead of and behind the sample plug are eliminated resulting in a partial purification of the sample based on mobility.

Cross injection also generates tighter sample plugs. During a direct injection, the sample plug becomes wider as the injection time progresses. In cross injection, the sample plug injects at a steady state flow resulting in a sample plug <1μm. Compared to direct injection, cross injection plugs can reduce the required channel lengths below twenty centimeters and still provide resolution comparable to forty-centimeter long capillaries.

Another favorable attribute of micro-channel devices is their superior heat dissipation that allows for even higher run voltages and faster separation times. The large thermal mass of the wafer allows for excellent heat dissipation from the small channels. Micro-channel devices typically run at 300 volts/cm while capillary systems operate at <140 volts/cm.

Sample detection in micro channel devices(figure 1) is very similar to capillary devices. Samples are electrophoresed toward the center of the wafer where they are excited with an argon laser beam. The laser is focused through a rotating microscope objective that illuminates each channel about three times per second. The resulting fluorescence is passed back through the objective and is directed to various beam splitters and filters. This light is then detected by one of four photo multipliers that sends the information to a computer for analysis.

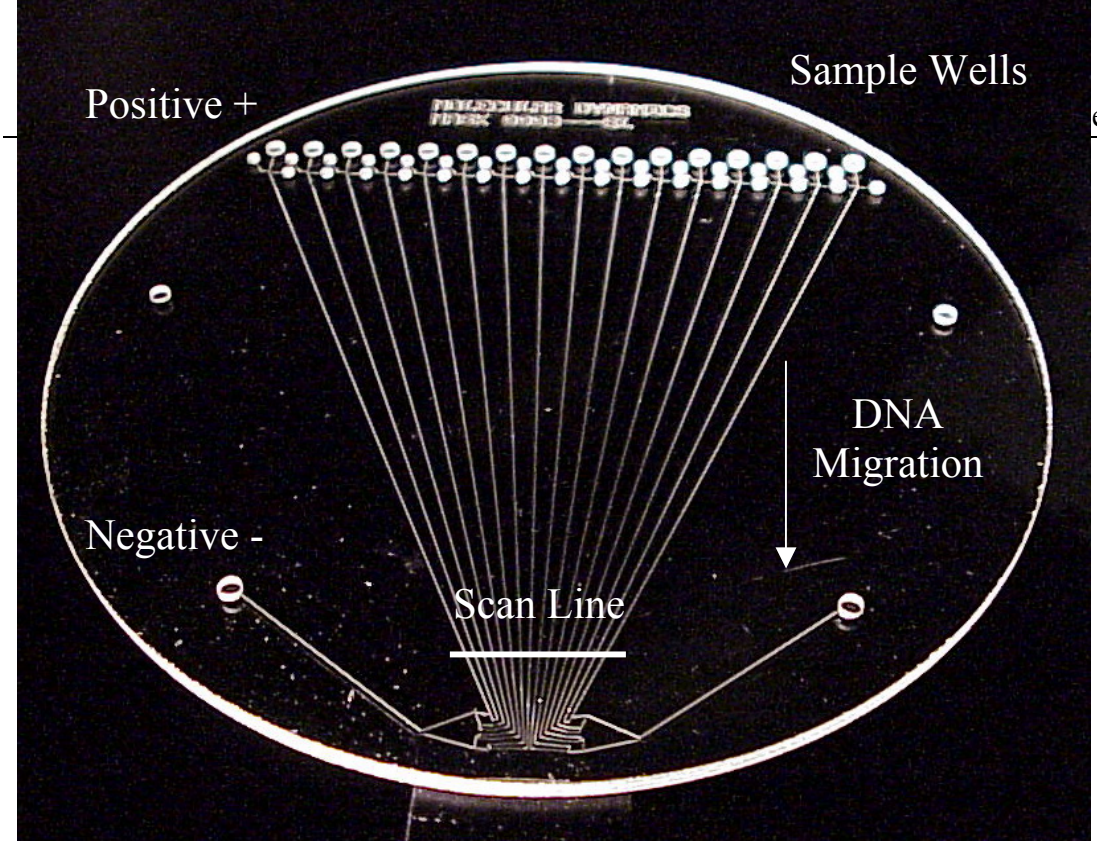

Figure 1 Micro-channel electrophoresis chip used for separation and detection of labeled DNA sequencing fragments. This 16-channel chip was fabricated at Molecular Dynamics in Sunnyvale California by Dr. Stevan Jovanovich

In the next decade, we may see micro-channel devices and high-density capillary arrays dominate high-throughput sequencing. However, these systems are limited to an increase in throughput of about four-fold over the MegaBACE 4000. More radical approaches that eliminate the time-consuming electrophoresis and sample preparation steps are needed to reach the ultimate goals of the genome initiative. The following sections briefly reviews two of the most promising technologies. Interested readers are directed to reviews that are more comprehensive.

Nanopore

The future of high-speed chemical analysis resides with novel sub-nanotech technologies combined with single molecule detection. These methods involve the difficult tasks of discerning structures less than a nanometer in diameter and the handling of femtoliter volumes. As these fields advance, most of the sample preparation and cost associated with DNA sequencing will disappear.

Nanopore technology or single channel current measurements is one of the newest methods being applied to DNA sequencing. It is based on the study of membrane ion channels that are formed by proteins in lipid membranes. By attaching a micrometer diameter glass pipette around the protein channel and placing electrodes on either side of the membrane, one can characterize the flow of ions through the channel by measuring the current flow.

The idea is to place DNA in the glass tip and measure the changes in ion current flow as the DNA transverses the pore. Measurements of current fluctuations in time intervals equal to a single base's translocation time may reveal the base's identity. Most of the current research uses the protein α-hemolysin and focuses on proof of principle and physical characterizations. α-hemolysin is a predictable membrane ion channel protein with pico-amp resolution. Its limiting pore diameter of 1.5 nanometers just allows single stranded DNA to experience some drag as it flows through the pore.

Recent studies have detected differences between poly-A and poly-C strands that are one hundred bases long. However, single base resolution is required for a feasible sequencing technology. Further studies have reveled that α-hemolysin's five nanometer long pore interacts with about seven DNA bases at one time. Since the current measurements would average all seven base interactions, single base resolution may not be possible with this pore. Modern current measuring equipment is also not accurate enough to measure the current fluctuations associated with single base interactions. While nanopore technology seems feasible, commercial instruments are at least a decade away.

Atomic Force Microscopy

Another interesting approach is the use atomic force microscopy (AFM) to directly read the sequence of a DNA strand. AFM uses a flexible cantilever attached to a nano size tip that is scanned over the sample surface. As the tip moves across the surface, the cantilever deflection is monitored by a laser that controls the tip height. This deflection angle is used to generate a topological map.

Using AFM to measure the force required to unzip DNA has given encouraging sequencing results. One research group attached one end of a double stranded DNA to the AFM tip and unzipped the DNA one base at a time by moving the tip. They found that A-T pairs required 10 piconewtons and G-C pairs required 20 piconewtons to unzip. Using a similar technique with glass needles, researchers have resolved 500 base pairs in 25 seconds. Although this is in early development, the technology may prove to be a valuable resource to specialized sequencing or genotyping applications.

CONCLUSION

The development of new technologies to automate, integrate and generally simplify the field of genomics and biotechnology is an ongoing process. Most large public genome centers are retooling their production lines and automation every 18-24 months to keep up with the improvements. Despite this rapid change, the sequencing phase of the human genome project has only been the start point of applying engineering and factory-like principals to bioprocesses. As the next phases of the genome project lead us to molecular medicine and point of care diagnostics, the

automation of genomic analysis will become even more central to the future of this field. The initial academic prototype development has grown into a vibrant support industry that is now developing technologies and making these available as discovery systems to academic and commercial sources. The next few years will see the merging and integration of much of the technologies described in this chapter. The concept of a single device to rapidly read out genetic information such as DNA sequence from a blood sample or other material in not far away.

BIBLIOGRAPHY

Suggested reading

Kheterpal I, Scherer JR, Clark SM, Radhakrishnan A, Ju J, Ginther CL, Sensabaugh GF, Mathies RA (1996) DNA sequencing using a four-color confocal fluorescence capillary array scanner. *Electrophoresis* **17:**1852-1859.

Zhou H, Miller AW, Sosic Z, Buchholz B, Barron AE, Kotler L, Karger BL (2000) DNA sequencing up to 1300 bases in two hours by capillary electrophoresis with mixed replaceable linear polyacrylamide solutions. *Analytical Chemistry* **72:** 1045-1052.

Medintz IL, Paegel BM, Mathies RA (2001) Microfabricated capillary array electrophoresis DNA analysis systems. *Journal of Chromatography A.* **924:** 265-70.

Paegel BM, Hutt LD, Simpson PC, Mathies RA (2000) Turn geometry for minimizing band broadening in microfabricated capillary electrophoresis channels. *Analytical Chemistry* **72:** 3030-3037.

Dolnik V, Liu S, Jovanovich S (2000) Capillary electrophoresis on microchip. *Electrophoresis* **21:** 41-54.

Deamer DW, Akeson M (2000) Nanopores and nucleic acids: prospects for ultrarapid sequencing. *Trends in Biotechnology* **18:** 147-51.

Rief M., Clausen-Schaumann H., Gaub HE (1999) Sequence-dependant mechanics of single DNA molecules. *Nature Structural Biology* **6:** 346-349.

Marziali A, Akeson M (2001) NEW DNA SEQUENCING METHODS. *Annual Review of Biomedical Engineering* **3:**195-223.

Meldrum D (2000) Automation for genomics, part two: sequencers, microarrays, and future trends. *Genome Research* **10:**1288-1303.

Schmalzing D, Koutny L, Salas-Solano O, Adourian A, Matsudaira P, Ehrlich D (1999) Recent developments in DNA sequencing by capillary and microdevice electrophoresis. *Electrophoresis* **20:** 3066-3077.

